

etimoalogik jihatdan “sarikor” shaklidagi qaratqich aniqlovchili forscha soʻz birikmasining izofali koʻrsatkichi -i ning tushib qolishi natijasida hosil qilingan qoʻshma otdir. Bu soʻz dastlab “sari kor” soʻz birikmasi shaklida “ishning boshi” kabi maʼnosida qoʻllanilgan, keyinchalik mazkur soʻz birikmasi tarkibidagi -i izofasi tushirib qoldirilgan va “sarkor” shaklidagi “ish boshi” kabi maʼnoni ifodalovchi qoʻshma sozga aylantirilgan. Hozirgi kunda ham xuddi shunday shakl va maʼnoda isteʼmolda qoʻllanilib kelinmoqda. Asar tilida yana shunga oʻxshash forscha qaratqich aniqlovchili soʻz birikmalaridan “oluyi Buxoriy” muruki Qarshi” kabi forscha soʻz birikmalaridan foydalanilgan. “Yana “ oluyi Buxoriy” mashhurdir. Buxoro olusidek hech yerda boʻlmas, terisini qurutuvchi tabarruklik bila viloyattin viloyatqa eltarlar” “Qarshi viloyatida behad va banihoyat boʻlur uchun ul navohida “murgʻaki Qarshi” derlar. Jumladan, “Boburnoma”da forscha asliy va nisbiy sifatlar ham ancha-munchani tasgkil etadi. Oʻzbek tili bilan fors tili sifatleri deyarli bir xildir. Oʻzbek asliy sifatleri daraja koʻrsatkichlarini qabul qilsa, fors tili asliy sifatleri ham xuddi shunday xususiyatlarga aga. Asarda “xoʻb” sifatidan juda unumli foydalangan. “Andijon toʻrt yigʻoch yoʻldir., havosi xoʻbva oqar suvlari farovondur.” “ Siyoq ilmini xoʻb bilur edi.” “Atrofida tort xoʻb ulang voqe boʻlubdir” Yuqoridagi asardan keltirilgan iqtiboslardan birinchisida boshqa gaplar singari “xoʻb” soʻzi qoʻllanilgan, bu soʻz asli forscha boʻlib, leksik jihatdan “yaxshi” soʻzi maʼnosini ifodalaydi. “YAXSHI” soʻzi oʻzbek tilida narsalarning belgi-xususiyatini bildiruvchi asliy sifatlardan biri sanaladi. Fors tilida ham “xoʻb” shunday xususiyatlarga ega boʻlgan asliy sifat sanaladi. Birinchi gapda qoʻllanilgan “xoʻb” soʻz turkumi jihatdan asliy sifat boʻlsa ham gap boʻlagi jihatdan kesim vazifasida kelgan. Keyingi gapdagi “xoʻb” soʻzi esa u ham soʻz turkumi jihatdan sifat boʻlsa ham gap boʻlagi sintaktik jihatdan hol vazifasida kelgan. Masalan, “ Siyoq ilmini xoʻb bilur edi.”. Mazmuni: “ Yozuv sanʼatini juda yaxshi bilar edi”. Demak, biz ushbu gapdagi “xoʻb bilur edi.”. jumlasiga “qay tarzda” kabi soʻroq beramiz. Navbatdagi iqtibosda esa “xoʻb oʻlang voqe boʻlubdur” kabi jumlasini ishlatilgan. Ushbu jumlada “xoʻb oʻlang” sifatlovchi aniqlovchili soʻz birikma boʻlib kelgan va bu soʻz “oʻlang – oʻtloq” soʻzi xususiyatlarini aniqlab kelmoqda. Mazkur gapdagi “ xoʻb” leksemasi turkum jihatdan ham gap boʻlagi jihatidan ham oʻz oʻrnida qoʻllanilgan. Keyingi iqtiboslarda ham “xoʻb” soʻzi gaplarda feʼllarga bogʻlanib kelib soʻz turkumi jihatdan sifat turkumida kelib, gap boʻlagi jihatdan holat ravishi maʼnosini ifodalagan.

Xulosa qilib shuni aytish mumkinki, Bobur oʻzining “Boburnoma” asari tilida forscha oʻzlashmalar, oʻzbek tili asliy va nisbiy sifatleri bilan bir qatorda ularga sinonim (maʼnodosh) forscha sifatleri, izofali birliklardan ham oʻz oʻrnida unumli foydalangan. Ular asar tilining chiroyli, rang-barang, taʼsirchan hamda xalqona boʻlishligini taʼminlagan. Asar tilini tilshunoslik nuqtayi nazarida tahlil etar ekamiz asarda forscha asliy sifatlardan 65 tasi qoʻllanilgan. Ularning ayrimlarinigina leksik- semantic va Grammatik tahlilini amalga oshirdik.

FOYDALANILGAN ADABIYOTLAR ROʻYXATI:

1. Abdurahmonov Gʻ. “Oʻzbek tilining tarixiy grammatikasi . – T; Oʻqituvchi. 1973.
2. Abdugʻafurov A. “Bobur sheʼriyati” Devon. T; Fan nashrilomyoti 1994. 79-bet.
3. Ibrohimov A. “ Boburnoma- buyuk asar” T; Gʻ.Gʻulom nashriyoti 2000. 321-bet.
4. Xolmanova Z. “Boburnoma” ning ayrim leksik xususiyatlari” T. Mirzo Ulugʻbek nomidagi Oʻzbek Milliy Universiteti. 2003. 99-bet.
5. Vafoyev.B “ Koʻhna soʻzlar tarixi” T. Fan. 1991. 103-bet.
6. Qudratillayev “Boburnoma” yangi nashr Oʻz AS 2003.

UDC 81'33

COMPILING A PARALLEL CORPUS FOR TRANSLATION RESEARCH

D.E.Inamova, senior teacher, Uzbekistan state university of world languages, Tashkent

M.A.Ismatova, teacher, Uzbekistan state university of world languages, Tashkent

Annotatsiya. Tarjima qilingan tarkibni bir necha sahifalardan tashqari tekshirish koʻplab omillar tufayli qiyin boʻlishi mumkin. Korpus lingvistikasi katta hajmdagi tarjima materiallarini ilgʻor tahlil qilish imkonini beradi. Parallel korpus tarjimaning yangi oʻrganilmagan jihatlarini sinchkovlik bilan oʻrganishni taʼminlaydi. Ushbu maqola parallel korpusni yaratish boʻyicha qaror

qabul qilish jarayonini qamrab oladi va parallel korpuslarga xos yaratish bilan bog'liq bo'lgan masalalariga e'tibor qaratadi.

Kalit so'zlar: korpus lingvistikasi, parallel, parallel korpus, manba tili, tarjimalar

Аннотация. Изучение переведенного контента, превышающего несколько страниц, может оказаться сложной задачей из-за множества факторов. Корпусная лингвистика позволяет проводить расширенный анализ больших объемов переведенного материала. Параллельные корпуса обеспечивают тщательное исследование новых, неизведанных аспектов перевода. В этой статье рассматривается процесс принятия решений по созданию параллельного корпуса, уделяя особое внимание проблемам компиляции, характерным для параллельных корпусов.

Ключевые слова: корпусная лингвистика, параллель, параллельные корпуса, исходный язык, переводы.

Abstract. Examining translated content beyond a few pages can be challenging due to the numerous factors involved. Corpus linguistics allows for advanced analysis of large amounts of translated material. Parallel corpora provide careful research of newly unexplored aspects of translation. This article covers the decision-making process for creating a parallel corpus, focusing on compilation concerns specific to parallel corpora.

Key words: corpus linguistics, parallel, parallel corpora, source language, translations

The structure of parallel corpus. A parallel corpus is typically structured in the form of aligned parallel texts, where each text in one language is aligned with its corresponding translation in another language. The alignment can be at the sentence level or at a more granular level, such as at the word or phrase level.

In a parallel corpus, each pair of aligned texts is typically stored in a parallel text unit, with metadata about the source and target languages, as well as any other relevant information such as the domain or genre of the texts.

Parallel corpora are often stored in a format that allows for easy retrieval and processing, such as in a database or in a structured file format such as XML. This structure allows for efficient searching and analysis of the parallel texts, as well as for the development of tools and applications that use the parallel corpus data.

Some parallel corpora also include additional linguistic information, such as part-of-speech tags or syntactic annotations, which can be used for more advanced linguistic analysis and machine learning applications. A corpus is essentially a huge yet systematic collection of naturally occurring, authentic texts maintained in digital form. A parallel corpus combines source materials in one language (L1) with translations into another language (L2). Parallel concordances enable simultaneous examination of both source texts and translations.



Picture 1. Structure of a Unidirectional Parallel Corpus

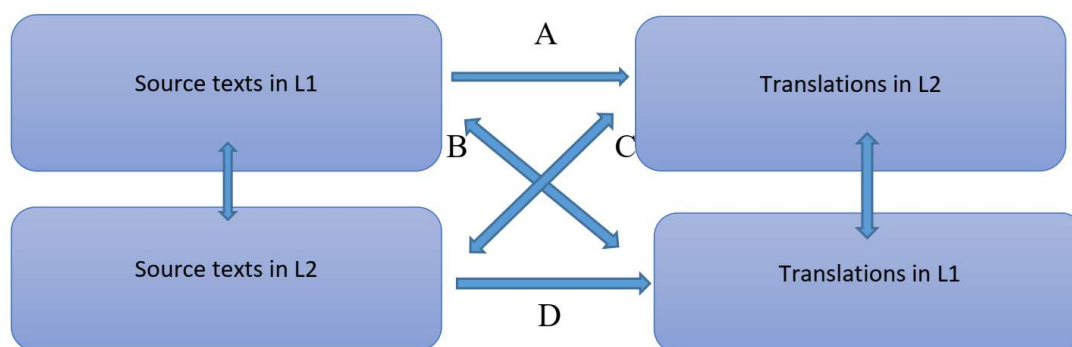
Parallel corpora can be classified as unidirectional, bidirectional, or a combination of both. Figure 1 illustrates the unidirectional arrangement, which uses source texts in L1 and translations into L2. A bidirectional corpus consists of source texts in two languages (L1 and L2) and their corresponding translations into L1 and L2.



Picture 2. Structure of a Bidirectional Parallel Corpus

A bidirectional structure allows researchers to analyze translations from L1 into L2 and L2 into L1, as seen in Picture 2. Parallel corpora with a mixed structure include both unidirectional and bidirectional arrangements.

By aligning unidirectional parallel corpora, lexicographers and linguists can create or improve bilingual dictionaries, computer lexicons, and grammars. Translators and students can use these tools to analyze how words or text segments are translated.



Picture 3. Possible Directions of Analysis in a Bidirectional Parallel Corpus

The L1-L2 and L2-L1 configurations of bidirectional parallel corpora enable a variety of different investigations. Picture 3 demonstrates a bidirectional structure that allows researchers to conduct multilingual investigations from L1 to L2 and L2 to L1 (arrows A and B). Bidirectionality is crucial because translations across languages are not always clear. For example, translating X into Y does not guarantee that Y will be translated as X. Creating an L2-L1 lexicon requires more than just reversing an L1-L2 dictionary, as experienced translators understand. To completely comprehend L1-L2 correspondences and mismatches, each language directions must be evaluated individually, notwithstanding potential crossovers. The bidirectional nature of a parallel corpus may be used in investigations using comparable corpora of translated and non-translated texts in the same language (arrows C and D in Picture 3). Comparing translated and non-translated texts can reveal unique characteristics of translation, such as L1 (source texts) vs. L2 (translations).

Compiling Parallel Corporations. Before creating a parallel corpus, several decisions must be made in addition to those for a monolingual corpus, such as genres and language varieties to include, whether to include spoken or written texts, and whether to use older or more recent texts. To begin, determine which language pairings to represent and their unidirectional or bidirectional relationship in the corpus. The corpus should comprise professional, learner, published, and native speaker translations. Combining these options might be complex and limited by the available translated texts. Only a small portion of people's writing is translated, limiting the quantity of texts accessible for parallel corpora.

Parallel corpora often have smaller sizes than monolingual corpora. Another thing to keep in mind is that some language pairings are more common than others. Although there are many English-Uzbek translations, few materials are translated from Russian to Uzbek.

There is a gap in translation availability across different text types, genres, forms, and time phases. Translations of spontaneous speech are uncommon, unlike film subtitles, which are commonly available. Despite having a large quantity of translated writings in a certain topic, most are unidirectional. Many films, videos, and DVDs are translated from English to Uzbek, but few from Uzbek to English.

When creating a corpus, it's important to decide if it will be utilized privately by a small group or publicly. Obtaining copyright authorization to utilize public corpora may be time-consuming and difficult, despite their ability to be shared and checked.

It's important to note that parallel corpora require double permissions. Obtaining copyright clearance for a source work does not guarantee authorization to use the translation. Even when using public domain source texts, translations may still be subject to copyright laws.

Compiling parallel corpora involves gathering and aligning text data in two or more languages for the purpose of creating a parallel corpus. This process typically involves the following steps:

1. Identifying the languages of interest: Determine which languages you want to include in your parallel corpus. The corpus is made up of text extracts taken randomly from the beginning, middle

and end of books. Extracts were preferred over full texts in order to facilitate the procurement of copyright permissions for a corpus that was to be made available online.

2. Collecting text data: Gather text data in each language from various sources, such as websites, books, articles, and other written materials.

3. Aligning the text data: Align the text data in each language so that corresponding sentences or phrases are matched up with each other. This can be done manually or using automated alignment tools. COMPARA aligns based on source-text sentences, unlike other parallel corpora that do not differentiate between one-to-many and many-to-one alignment. When there was no one-to-one relationship between source and translation sentences, they were split or linked with nearby phrases to reflect the original structure.

4. Cleaning and preprocessing the data: Remove any noise or irrelevant information from the text data and preprocess it by tokenizing, lowercasing, and lemmatizing the text.

5. Creating the parallel corpus: Combine the aligned and preprocessed text data in a single file format that indicates the relationship between the texts in each language.

6. Verifying the alignment: Check the alignment of the texts to ensure that they are accurately matched and that there are no errors or inconsistencies.

To gain permission to use a text in a corpus, first determine who owns the rights, which may be writers, publishers, translators, or heirs of deceased writers and translators. One of the biggest challenges for corpus compilers is explaining what a corpus is to copyright holders. Many people are unfamiliar with corpora and fear that storing a copy of their texts on a server will result in illegal downloads. Copyright holders should be informed that users of a corpus typically obtain concordances and frequency lists, not entire texts. Limiting the quantity of material downloaded from a corpus is also an option. Copyright holders can be reassured that providing concordances through a corpus promotes the entire text and encourages consumers to purchase it. If you want to create a public corpus, you need first address copyright clearances.

By following these steps, you can compile a high-quality parallel corpus for translation research that can be used to train and evaluate translation systems, study cultural and linguistic differences, and improve machine translation algorithms. Overall, compiling parallel corpora is a time-consuming process that requires careful attention to detail and accuracy to ensure the quality of the parallel corpus.

REFERENCES:

1. Baker, M. 1993. Corpus linguistics and translation studies. Implications and applications. In M. Baker, G. Francis and E. TogniniBonelli (eds.) Text and Technology: In Honour of John Sinclair (pp 233-250).
2. Amsterdam: John Benjamins. Barlow, M. 2002. ParaConc: Concordance software for multilingual parallel corpora. In proceedings of LREC-2002: Third International Conference on Language Resources and Evaluation, (pp.20-24) Las Palmas.
3. Frankenberg-Garcia, A. & Santos, D. 2003. Introducing COMPARA, the Portuguese-English parallel translation corpus. In F. Zanettin, S. Bernardini & D. Stewart (eds.) Corpora in Translation Education, (pp. 71-87) Manchester: St. Jerome Publishing.

UO'K 81-11

INGLIZ VA O'ZBEK TILLARIDA OILA KONSEPTINING CHOG'ISHTIRMA TAHLILI

H.Ishqobilova, o'qituvchi, O'zbekiston davlat Jahon tillati universiteti, Toshkent

Annotatsiya. Ushbu maqolada ingliz va o'zbek tillarida milliy va ma'naviy qadriyat sifatida ifodalanib keluvchi oila konsepsiyasi ingliz tilidagi (*Love is an everyday thing*) va o'zbek tilidagi (*Er libosli ayol*) asarlarda tahlil qilingan va qanday ma'nolarni ifoda etishi ko'rib chiqilgan. Ikki tildagi asarlar tahlili jarayonida turli xalq vakillari tomonidan oila konseptini turli vaziyatlarda qo'llash uchun tanlangan leksemalarning o'xshash va farqli jihatlari ko'rib chiqilgan. Quyidagi tahlil natijasi shundan dalolat beradiki, oila atamasining asarda turli xil ma'nolarda qo'llanilishi mazkur tilda bu